

***Measuring Uncertainty by Extracting
Fuzzy Rules Using Rough Sets***

Jeffrey A. Worm

&

***Extracting Fuzzy Rules Under
Uncertainty and Measuring
Definability Using Rough Sets***

Donald E. Culas

RICIS Preface

This research was conducted under auspices of the Research Institute for Computing and Information Systems by Jeffrey A. Worm and Donald E. Culas of the University of Houston - Downtown. Dr. Andre' de Korvin was the UH - Downtown faculty advisor. Dr. A. Glen Houston served as the RICIS research coordinator.

Funding was provided by the Information Systems Directorate, NASA/JSC through Cooperative Agreement NCC 9-16 between the NASA Johnson Space Center and the University of Houston-Clear Lake. The NASA technical monitor for this activity was Dr. Robert Savely of the Information Technology Division, NASA/JSC.

The views and conclusions contained in this report are those of the author and should not be interpreted as representative of the official policies, either express or implied, of UHCL, RICIS, NASA or the United States Government.

MEASURING UNCERTAINTY BY EXTRACTING FUZZY RULES USING ROUGH SETS

Prepared by
Jeffrey A. Worm

in
Partial Fulfillment
of the Requirements
for
CS 4395

Department of Applied Mathematical Sciences
University of Houston - Downtown
November 20, 1991

Committee Members / Approval

<i>Dr. Andre' de Korvin</i>	Faculty Advisor	_____
<i>Dr. Kenneth Oberhoff</i>	Member	_____
<i>Dr. Ananda Gunawardena</i>	Member	_____
<i>Dr. Robert Lea</i>	Member	_____
<i>Dr. Richard A. Alo'</i>	Chairman	_____

ORIGINAL PAGE IS
OF POOR QUALITY

MEASURING UNCERTAINTY BY EXTRACTING FUZZY RULES USING ROUGH SETS

TABLE OF CONTENTS

	Page
Abstract	
Introduction	2
Section 1: Uncertainty	6
1.1 Uncertainty.....	6
1.2 Techniques to Combat Uncertainty.....	7
1.2.1 Statistics.....	8
1.2.2 Dempster-Shafer Theory.....	9
1.2.3 Fuzzy Set Theory.....	10
1.2.4 Rough Sets.....	12
1.3 The Proposed Solution.....	14
Section 2: Rough Set theory vs. Fuzzy Set theory	16
2.1 Rough Sets - a closer examination.....	16
2.2 Fuzzy Sets - a closer examination.....	19
2.2.1 Functions of Fuzzy Set Properties.....	20
2.3 Establishing Certain and Possible Rules.....	21
2.3.1 Threshold Values.....	23
2.3.2 Extracting Possible and Certain Rules.....	24
Section 3: Application: Tumor Diagnosis	27
Section 4: Software Design and Specifications	31
4.1 Software Specifications.....	31
4.1.1 Input.....	32
4.1.2 Functions and Calculations.....	33
4.1.3 Output.....	34
References	35

Abstract

Despite the advancements in the computer industry in the past thirty years, there is still one major deficiency. Computers are not designed to handle terms where uncertainty is present. To deal with uncertainty, techniques other than classical logic must be developed. This paper examines the concepts of statistical analysis, the Dempster-Shafer theory, rough set theory, and fuzzy set theory to solve this problem. The fundamentals of these theories are combined to provide the possible optimal solution. By incorporating principles from these theories, a decision-making process may be simulated by extracting two sets of fuzzy rules: certain rules and possible rules. From these rules a corresponding measure of how much we believe these rules is constructed. From this, the idea of how much a fuzzy diagnosis is definable in terms of its fuzzy attributes is studied.

Acknowledgements

I would like to express my thanks to Dr. Andre' de Korvin for his very helpful advice throughout the semester. It is his paper, *Extracting Fuzzy Rules under Uncertainty and Measuring Definability using Rough Sets*, which provided the cornerstone for my research and the blueprint for my software.

I would also like to extend my thanks to Dr. Kenneth Oberhoff for his helpful insights in all phases of software design.

Finally, I also extend my thanks to Dr. Richard A. Alo' for his constructive comments throughout the semester regarding presentations and all initial drafts of this document.

INTRODUCTION

Computers have progressed so much over the past thirty years that it is now hard to imagine life without them. They have become smaller, faster, and less expensive. Similarly, the applications we use them for has grown exponentially. If the auto industry had done what the computer industry has done in this time, a Rolls-Royce would cost a couple of dollars and get a couple of million miles per gallon.

An important development of this progression is the computer's ability to refine and expedite the decision-making process. One can enter in raw data as input and receive the output in an organized, logical form. This manipulated form may then be used to help facilitate some type of decision by the user. It is also possible for a computer program to have a built in "thinking" function which requires no help from the user in order to formulate a decision. A decision may be automatically made by the computer, solely on the output and any preset conditions of the output.

A program which can perform these simple functions is possible

through knowledge acquisitions using examples. Through repetition, one may learn to associate certain factors to form a decision. Ideally, the decisions will always be the same if the corresponding factors are always the same. For example, if a person sees lightning and hears thunder, they may assume it is raining close by from some similar experiences in the past. Again, this is under "ideal" circumstances; the person is positive they see lightning and positive they hear thunder. Unfortunately, "ideal" circumstances are not always present.

As amazing as the progression of computers has been, there is a noticeable deficiency: computers are not designed to manipulate data where uncertainty is present. Uncertainty may arise in many different ways. It may be brought about by ambiguous terms used to describe a certain situation. It may also be caused by scepticism of rules used to describe a cause of action, or by missing or erroneous data. To handle uncertainty, methods other than classical logic must be developed. One possible solution to this is to use fuzzy set theory to extract rules.

In ordinary set theory, an element is either in or out of the set. In **fuzzy set** theory, however, an approximation is used to determine the degree to which an element is in the set. This is due to the fact that subjective terms are often used to describe a condition. Fuzzy set theory

allows for a fraction of an element to be in the set. From these fuzzy sets, one can extract two sets of fuzzy rules: certain rules and possible rules. Basically, the certain rules are formed by taking the minimum of the union of the fuzzy sets. Conversely, the possible rules are formed by taking the maximum of the intersection of the fuzzy sets.

Another possible solution to deal with uncertainty is in learning from examples. An effective method to acquire knowledge through examples is rough sets. **Rough sets** is the theory of endorsements and non-functional logic. As in fuzzy set theory, possible and certain rules are extracted. In rough sets, these rules are generated by qualities known as the upper and lower approximations. These qualities are similar to the inner and outer reductions of Dempster-Shafer theory. The attributes of the conditions are assigned values and a measure of how much these attributes determine the diagnosis is established. However, the values of these attributes require some judgement for their determination. Similarly, the diagnosis is often not of "pure" type, but a combination which is reflective of fuzzy sets.

Combining these two methods of fuzzy set and rough set theories, as well as the principles of Dempster-Shafer theory, provides a possible optimal solution for dealing with uncertainty. By integrating these two

methods, we can produce a set of certain rules and possible rules and determine a measure of belief associated with these rules. These rules allow a basis of dealing with uncertainty in the decision-making process.

- Section 1 -

Uncertainty

1.1 Uncertainty

The previously referenced computer program took certain variables and "crunched" them up to come to a certain decision. A major question that arises is, "How does one deal with uncertainty?". Uncertainty may arise in many different situations. It may be caused by the ambiguity in the terms used to describe a specific situation, or it may be caused by the skepticism of rules used to describe a course of action. Uncertainty may also be caused by inconsistencies in data, or simply by missing or erroneous data.

To understand what is meant by ambiguity of terms, one must realize that different people may associate different meanings or values for the same term(s). To illustrate this, one cannot put a set value on "very rich" or "moderately rich" because these are subjective terms. One person's definition may be quite different from another's. For this reason, descriptive terms may contain some degree of ambiguity, and therefore some degree of uncertainty.

Uncertainty caused by the skepticism of rules may be attributed to an underlying doubt one may have regarding a situation. Occasionally, all factors may point towards a certain decision, but one's "gut feeling" produces a degree of doubt toward that decision. Whether these doubts are warranted or not, they must be taken into account when we refer to uncertainty. For these doubts may influence one's future decisions on similar situations.

Clearly, any missing or erroneous data will lead to uncertainty. Unfortunately, it is not always obvious when data is wrong. A strong characteristic of erroneous data is inconsistencies. In other words, if the same data has conflicting outcomes, there is uncertainty present. To illustrate this, the table below represents how a decision-maker may make an inconsistent decision based on a couple of pieces of data. In this example, Case X_2 and Case X_5 have the same data, yet different decisions. This shows that uncertainty exists somewhere in this decision-making process.

<u>CASE</u>	<u>CONDITIONS</u>		<u>DECISION</u>
	<u>DATA1</u>	<u>DATA2</u>	
X_1	0	0	A
X_2	1	0	B
X_3	0	1	B
X_4	1	1	A
X_5	1	0	A

1.2 Techniques to Combat Uncertainty

1.2.1 Statistics

To deal with uncertainty, techniques other than classical logic need to be developed. The most useful tool for handling probability is statistics, or statistical analysis. **Statistical analysis** is concerned with the collection, organization, and interpretation of data according to well-defined procedures. Observations are made and converted into numerical form. The numbers are manipulated and organized, with the results interpreted and translated back into a way one may understand.

Statistical analysis allows for the reduction of data. Large masses of unorganized numbers may be characterized into smaller sets that describe the original observations without sacrificing critical information. The second major role lies in its use as an inferential measuring tool. In other words, it provides procedures for stating the degree of confidence one may have in the accuracy of the measurements one makes. Finally, statistical analysis allows one to make distinctions about relationships that exist between and among sets of observations. Does knowledge about one set of data allow us to infer or predict characteristics about another set of data?

Statistical analysis does, however, have some deficiencies. Data

reduction may lead to the sacrificing of detail. The inferential measuring tool statistical analysis provides is useful, but all measurements are subject to error. Furthermore, sometimes one may strive to find a connection between two sets so much that a connection is unjustifiably made.

Though statistics is a useful method for handling probability, it provides only a foundation for the problem of knowledge acquisition under uncertainty. Three theories which are better suited to handle this problem are: Dempster-Shafer Theory, fuzzy set theory, and rough set theory.

1.2.2 Dempster-Shafer Theory

The Dempster-Shafer Theory is a theory of evidence and probable reasoning. It is a theory of evidence because it deals with weights of evidence and with numerical degrees of support based on evidence. It is a theory of probable reasoning because it focuses on the combination of evidence, more specifically, the combination of belief functions.

The theory begins with the idea of using a number between zero and one to indicate the degree of belief one should assign for inclusion on the basis of the evidence. Its focus lies in the combination of degrees of belief based on one body of evidence with those based on an entirely

distinct body of evidence. This combination of belief functions is the heart of the Dempster-Shafer Theory. Given several belief functions based on distinct bodies of evidence, this theory enables one to compute a new belief function based on the combined evidence.

The main connection Dempster-Shafer has to the other theories is the concepts of inner and outer reductions. As will be shown in the discussion concerning rough sets (section 2.1), this concept is almost identical to the lower and upper approximations of rough sets. In inner reduction, denoted by $\theta(A)$, is the largest subset that implies A. The outer reduction, denoted by $\overline{\theta}(A)$, is the smallest subset that is implied by A.

This theory is very similar to rough set theory.

1.2.3 Fuzzy Set Theory

Perhaps the most useful tool when dealing with uncertainty is fuzzy set theory. This theory is the most practical where ambiguous terms are present. To get a complete understanding of this theory, one must first backtrack to ordinary set theory. All branches of mathematics are developed, consciously or unconsciously, in set theory or some part of it. It is, therefore, an important concept to grasp. A set is a collection of things (called elements or members), the collection being regarded as a single object. An item is either in the set or it is not. This property is

referred to as inclusion.

In **fuzzy set** theory, however, an approximation is used to determine the degree to which an element is in the set. Such concepts as inclusion or set equality may seem too strict. Usually, the structures embedded in fuzzy set theories are less rich than the boolean lattice of ordinary set theory. Unlike ordinary set theory, one cannot determine the cardinality, or size, in fuzzy set theory. One cannot compute an accurate union or intersection of two fuzzy sets because the elements are estimates of inclusion, not "crisp" values.

If the value of a set is allowed to be the real interval $[0,1]$, A is called a fuzzy set. The grade of membership of an element, x , in A is $\mu_A(x)$. The closer the value of $\mu_A(x)$ is to 1, the more x belongs to A . Similarly, the lower the value of $\mu_A(x)$, the less x belongs to A . Clearly, A is a subset of x that has no crisp boundary. By using fuzzy set theory, one must approximate the value of inclusion an element has in a set.

Earlier the question was raised, "What is the difference between 'very rich' and 'moderately rich'?". Fuzzy set theory could approximate a person worth $\$X$ to be .4/very rich and .8/moderately rich. Because of the ambiguity of the term "rich", one needs to approximate the value of the person for "very rich" and "moderately rich". It might be observed that, for

the decision-maker assigning the values, the person falls into the category of "moderately rich" more than "very rich". For this reason, the decision-maker puts more "weight" on the term "moderately rich". The person lies within the set of "moderately rich" more than the set of "very rich". Hence, they are assigned those corresponding values.

A problem one may encounter using this theory is the fact that the decision-maker assigns these values. Obviously, not all people have the same pre-conceived meanings for terms such as "very rich" or "extremely tall". The approximations one person gives may be completely different from the approximations of someone else. For example, a small boy may see a man 5'9" as "very" tall. Conversely, a professional basketball player might see the same person as "average" height. It is best to keep this in mind, because it can easily influence the decisions.

1.2.4 Rough Sets

As was stated earlier, the most traditional way of acquiring knowledge is based on learning from examples. Another effective tool of inferring knowledge from examples is rough sets. **Rough sets** is the theory of endorsements and non-functional logic.

Let U be a non-empty set, call the universe, and let R be an equivalence relation on U , called an indiscernibility relation. An ordered

pair $A = (U, R)$ is called an approximation space. For an element x of U , the equivalence class of R containing x will be denoted by $[x]_R$. Equivalence classes of R are called elementary sets in A . We assume that the empty set is also elementary. Any finite union of elementary sets in A is called a definable set in A .

Two more concepts, known as the lower approximation and upper approximation of X in A are examined later. Basically, the lower approximation of X in A is the greatest definable set in A , contained in X . The upper approximation of X in A is the least definable set in A containing X . These concepts correspond to the inner and outer reductions from Dempster-Shafer Theory, also examined later. A rough set in A is the family of all subsets of U having the same lower and upper approximations in A .

There are essential connections between rough set theory and Dempster-Shafer theory. For example, the lower and upper approximations of rough set theory exist under the names of inner and outer reductions, respectively. Similarly, the qualities of lower and upper approximations of rough set theory are the belief and plausibility functions, respectively, of Dempster-Shafer theory.

The main difference between rough set theory and the Dempster-

Shafer Theory is in the emphasis: Dempster-Shafer Theory uses belief functions as a main tool, while rough set theory makes use of the family of all sets with common lower and upper approximations. The main advantage of rough set theory is that it does not need any preliminary or additional information about data.

1.3 The Proposed Solution

The main purpose of this work is to study the setting described before where a decision-maker is faced with uncertain (i.e. fuzzy) conditions and makes a fuzzy decision which might be strongly or weakly based on these symptoms. Here, the techniques of fuzzy set theory and rough sets will be incorporated to attempt to provide the optimal solution of measuring uncertainty. From the conditions and decisions, one will find that fuzzy rules may be extracted. In fact, one may extract two sets of rules: certain rules and possible rules. One may also determine a measure of how much they believe in these rules.

The main body of this work is examined in detail in Section 2. The basic notations and results necessary to fully understand these concepts are discussed here. Section 3 offers a detailed example of these concepts at work. It provides an everyday application, as well as an opportunity to

see how these principles are incorporated. The software requirements and design of this product, which simulates the basic ideas set forth here, is available in Section 4. The coding of this program may be found in Appendix A.

- SECTION 2 -

Rough Set theory vs. Fuzzy Set theory

As stated in Section 1, I believe the optimal solution for knowledge acquisition under uncertainty lies within the combination of fuzzy set and rough set theories. By integrating the fundamentals of these theories, I hope to measure and, where possible, minimize the degree of uncertainty. To best understand how the concepts of fuzzy sets and rough sets are to be incorporated, it is important to first grasp the main principles of rough sets.

2.1 Rough Sets - A Closer Examination

Let U be the universe, R an equivalence relation on U , and X any subset of U . If $[X]$ denotes the equivalence class of X relative to R , we can then define the foundation of rough sets. This is called the upper and lower approximations of X and is denoted, respectively, by:

$$R(X) = \{ X \in U / [X] \subset X \} \quad \text{and}$$

$$B(X) = \{ X \in U / [X] \cap X \neq \emptyset \}.$$

Once again, rough sets are the family of all subsets in U having the same

upper and lower approximations.

To examine these upper and lower approximations closer, we define an information system as the quadruple (U, Q, V, τ) where $Q = C \cup D$ and $C \cap D = \emptyset$. The set C stands for the set of conditions, and D is the set of decisions. We assume that C is equal to the set of attributes, Q . The set V stands for value and τ is a function from $U \times Q$ into V where $\tau(u, q)$ denotes the value of attribute q for element u . For example, the pulse rate q of patient u . The set C produces an equivalence on U by partitioning U into sets over which all attributes are constant. A rough set is classified by properties of its lower and upper approximations. The set is called roughly C -definable if:

$$B(X) \neq \emptyset \text{ and } R(X) \neq U.$$

The set is internally C -undefinable if:

$$B(X) = \emptyset \text{ and } R(X) \neq U.$$

The set is externally C -undefinable if:

$$B(X) \neq \emptyset \text{ and } R(X) = U.$$

The set is totally C -undefinable if:

$$B(X) = \emptyset \text{ and } R(X) = U.$$

To illustrate this, the table previously referenced is examined again.

For this example, Decision 'A' denotes sickness. The conditions produce a partition on $\{X_1, X_2, X_3, X_4, X_5\}$, namely $\{ \{X_1, X_5\}, \{X_2, X_3\}, \{X_4\} \}$. The decision-maker defines "sick" people by $X = \{X_2, X_4\}$. Thus, the lower and upper

<u>CASE</u>	<u>CONDITIONS</u>		<u>DECISION</u>
	<u>DATA1</u>	<u>DATA2</u>	
X_1	0	0	A
X_2	1	0	B
X_3	0	1	B
X_4	1	1	A
X_5	1	0	A

approximations are $\underline{B}(X) = \{X_4\}$ and $R(X) = \{X_2, X_3, X_4\}$. In this example,

$\underline{B}(X) \neq \emptyset$ and $R(X) \neq U$, therefore X is roughly C-definable. For an internally

C-undefinable set X in S we can not say with certainty that any $x \in U$ is a

member of X . For an externally C-undefinable set X in S we can not

exclude any $x \in U$ being possibly a member of X . This is similar to

Dempster-Shafer belief and plausibility functions of rough sets.

The difference between the lower and upper approximations may be attributed to the presence of inconsistencies. If it were not for the inconsistencies, the decision-maker's opinion would be in line with the upper and lower approximations produced by C . It is this difference

between $R(X)$ and $\underline{R}(X)$ that offers a measure of how well the diagnosis of the decision-maker follows the conditions. If the decision-maker is an "expert", the difference between the lower and upper approximations gives one a measure of how good conditions C are to determine the diagnosis. In other words, the more we trust the decision-maker, the more we believe how the conditions determine the diagnosis. Moreover, it is these lower and upper approximations which generate the rules that will be used as the basis for the decision-making process. These generated rules, called the certain and possible rules, will be examined closer in Section 2.2.

Unfortunately, there may be uncertainty in the conditions, as well as the diagnosis. The conditions and the diagnosis rarely partition the universe into "crisp" sets. This is due to the fact that most of the values of attributes are descriptive, and thus subjective terms. It is this that leads to the "fuzziness" of the conditions/diagnosis when trying to define the terms. This "fuzziness" can lead to overlapping, therefore rendering crisp partitions nearly impossible. At best one hopes the terms can be partitioned with as little overlapping as possible.

2.2 Fuzzy Sets - A Closer Examination

For all decision-making processes, it is the rules which guides one

towards a decision. Decision-making under uncertainty is no different. The problem lies within determining these rules. As stated in Section 2.1, the upper and lower approximations generate possible and certain rules. It is Fuzzy Set theory which allows one to extract these rules.

2.2.1 Functions of Fuzzy Set Properties

To understand how these rules are extracted, one must first be familiar with the notation. A fuzzy subset A of U is defined by the function:

$$\mu_A : U \rightarrow [0,1].$$

This simply states that the values of the fuzzy subset A fall between 0 and 1. If A and B are fuzzy subsets, the properties $A \cap B$, $A \cup B$, and $\neg A$ are defined by the functions: $\text{Min}\{\mu_A(x), \mu_B(x)\}$, $\text{Max}\{\mu_A(x), \mu_B(x)\}$, and $1 - \mu_A(x)$, respectively. The property $\neg A \cup B$ corresponds to the function $\text{Max}\{1 - \mu_A(x), \mu_B(x)\}$. These computed values are the foundation for extracting the rules. Therefore, it is very important to understand what is meant by the notation.

The first function, $\text{Min}\{\mu_A(x), \mu_B(x)\}$, is computed by matching up the corresponding elements of the fuzzy subsets and taking the minimum (in value) of the two. For example, given the two fuzzy subsets:

$$\begin{aligned} A &= (.3, .4, .7, .8, .6, .1) \quad \text{and} \\ B &= (.6, .2, .4, .3, .5, .4) \end{aligned}$$

One can compute $\text{Min}(A,B) = (.3, .2, .4, .3, .5, .1)$. The second function, $\text{Max}\{\mu_A(x), \mu_B(x)\}$, is similar in computation to the first. Instead of taking the minimum of the two, one takes the maximum, or greatest in value. Using the two previous subsets of A and B, one can compute $\text{Max}(A,B) = (.6, .4, .7, .8, .6, .4)$. The third function, $1 - \mu_A(x)$, is computed by taking one(1) minus the values of the fuzzy subset. Again, using the previous subset A, one can compute $1-A = (.7, .6, .3, .2, .4, .9)$. The last function, $\text{Max}\{1-A(x), B(x)\}$, is simply a combination of the second and third functions. First, one computes $1-A(x)$ then compares that to $B(x)$, taking the maximum of the two. For example,

$$\begin{array}{rcl} \text{Max}\{1-A, B\} & = & \text{Max}\{(.7, .6, .3, .2, .4, .9), (.6, .2, .4, .3, .5, .4)\} \\ \downarrow & & \downarrow \qquad \qquad \downarrow \\ \text{Max}\{1-A, B\} & = & (.7, .6, .4, .3, .5, .9). \end{array}$$

2.3 Establishing Certain and Possible Rules

Now that the fundamental properties (and corresponding notation) have been explained, we can define two functions of major importance to this work. These two functions are on pairs of fuzzy sets and allow us to extract the rules. We assume here that A and B denote fuzzy subsets of the same universe. The function $I(A \subset B)$ measures the degree to which A is included in B. This function computes the rules generated by certainty

and is defined as:

$$I(A \subset B) = \inf_x \text{Max}\{1-A(x), B(x)\}.$$

The function $J(A \# B)$ measures the degree to which A intersects B . This function computes the rules generated by possibility and is defined by:

$$J(A \# B) = \max_x \text{Min}\{A(x), B(x)\}.$$

The function $I(A \subset B)$ is computed by first finding $\text{Max}\{1-A(x), B(x)\}$, then taking the minimum term. For the previous fuzzy subset examples of A and B , we found the $\text{Max}\{1-A, B\} = (.7, .6, .4, .3, .5, .9)$. Since the minimum term is $.3$, $I(A \subset B) = .3$. The function $J(A \# B)$ is computed by first finding $\text{Min}\{A(x), B(x)\}$, then taking the greatest (in value) term. Again, using A and B we found $\text{Min}\{A, B\} = (.3, .2, .4, .3, .5, .1)$. Since the maximum term is $.5$, $J(A \# B) = .5$.

For the example used in this work, we assume the decision-maker is faced with different conditions, or attributes, and makes a decision based on the values of these attributes. To provide a more concise explanation of this work, we will limit the number of possible decisions to two (2). Similarly, we will limit the description an attribute may have to two (2). For example, size can only be measured as a degree of large and small. These limitations are made to explain when to compute the $I(A \subset B)$ and

$J(A\#B)$ values.

For the functions of $I(A \subset B)$ and $J(A\#B)$, A denotes the descriptions of the attributes, while B denotes the possible decisions. For each description, we must measure the degree to which it is included in decision 'A' as well as in decision 'B'. In addition to this, we also measure the degrees of intersections of the descriptions for each decision. For example, if we have attribute-1 with descriptions of 'W' and 'X', attribute-2 with descriptions of 'Y' and 'Z', and possible decisions of 'A' and 'B', we would need to compute all of the following:

$I(W \subset A)$	$I(Y \subset A)$	$I(W \cap Y \subset A)$
$I(W \subset B)$	$I(Y \subset B)$	$I(W \cap Y \subset B)$
$I(X \subset A)$	$I(Z \subset A)$	$I(X \cap Y \subset A)$
$I(X \subset B)$	$I(Z \subset B)$	$I(X \cap Y \subset B)$
$J(W\#A)$	$J(Y\#A)$	$J(W \cap Y\#A)$
$J(W\#B)$	$J(Y\#B)$	$J(W \cap Y\#B)$
$J(X\#A)$	$J(Z\#A)$	$J(X \cap Y\#A)$
$J(X\#B)$	$J(Z\#B)$	$J(X \cap Y\#B)$

2.3.1 Threshold Values

As one can see, this leads to large numbers of rules. For this case, we would have 24 rules: 12 certain rules and 12 possible rules. If we had 3 attributes with 2 descriptions each, the number of rules would increase to 88 rules. It is therefore essential to establish a "threshold" value, denoted by α , for which we may ignore all rules falling below this value.

Actually, we need two of these values: one for the certain rules and one for the possible rules. The decision-maker may or may not set these two equal. The higher we set the threshold, the higher the belief we have for the rules which factor above it. Unfortunately, there is a trade-off, for the higher the threshold, the more rules we ignore. Ideally, the solution to this trade-off is to allow the decision-maker to interactively change the threshold values as they see fit. By allowing this interactive changing, it also provides somewhat of a sensitivity analysis. The decision-maker can immediately see which rules are affected by the changing threshold value. Another reason to promote interactive changing of the threshold is that the value of α is very much problem dependent. A value of $\alpha = .5$ might be the best for one problem, but irrelevant for another. The decision-maker may adjust the value till it is set at the most appropriate level.

2.3.2 Extracting Possible and Certain Rules

Once the threshold value has been established, it is time to extract the rules. All rules (values of I and J) which fall below the threshold value are immediately eliminated. To further eliminate rules, we have certain provisions. First, all rules with unique I and J values are kept. Second, if more than one rule has identical I values, we keep (extract) the "smaller" in terms of attributes. For example, if we were to obtain the

following certain rules:

If **W** then **A** is present .6 (1)

If **W** and **Y** then **A** is present .6 (2)

If **W** and **Z** then **A** is present .6 (3)

we would keep rule (1) because rules (2) and (3) offer no significant data.

Conversely, if these three rules were computed using J values, thus making them possible rules, we would extract rules (2) and (3). This is because rules (2) and (3) imply the possibility of rule (1).

The concepts discussed up to this point are represented through an example in section 3.

2.4 Definability of Terms

Now that all the certain and possible rules are extracted, we can measure the definability of terms. The goal of this is to define the terms in the decisions as a function of the terms in the conditions. How well this can be accomplished is a function of how much the decision follows the conditions.

Let $\{Q_i\}$ be a finite family of fuzzy sets. This family of sets does not necessarily form a partition on the universal set. Let A be a fuzzy set. A lower approximation of A through $\{Q_i\}$, produces the fuzzy set:

$$\underline{R}(A) = \bigcup \{Q_i \mid Q_i \subset A\}$$

Here, \cup denotes the union of fuzzy sets, and $I(Q_i \subset A)$ Q_i denotes the fuzzy set obtained by multiplying the components of Q_i by $I(Q_i \subset A)$. Therefore, if Q_i is very much a subset of A , $I(Q_i \subset A)$ Q_i is close to the whole set Q_i . Conversely, if $I(Q_i \subset A)$ is small, so is the contribution of Q_i to $\underline{R}(A)$. Similarly, we can define an upper approximation of A through $\{Q_i\}$ by:

$$\overline{R}(A) = \cup_j I(Q_j \not\subset A) Q_j.$$

In the special cases where all the sets are crisp, and $\{Q_i\}$ denotes a partition generated by an equivalence relation R , then the lower approximation is defined as:

$$\underline{R}(A) = \{X / [X] \subset A\},$$

and the upper approximation is defined as:

$$\overline{R}(A) = \{X / [X] \cap A \neq \emptyset\}.$$

One can therefore see that in this crisp case:

$$\underline{R}(A) \subset A \subset \overline{R}(A).$$

One should not, however, expect these inclusions to hold in the fuzzy case because boundaries of the relevant sets are poorly-defined.

- SECTION 3 -

Application: Tumor Diagnosis

As stated earlier in this paper, knowledge aquisition is best accomplished by looking at examples. It is therefore important to provide an example of the concepts discussed in section 2. By examining an application, these concepts should become clearer.

The analogy to be used here is that of a doctor (decision-maker) examining the characteristics (attributes) of tumors and rendering a diagnosis (decision). For this example, the attributes are size and color. Size can be described as large and small. Color will be limited to the descriptions of blue and red. The possible diagnoses will be either Disease 'D_A' and 'D_B'.

While examining seven patients, the following data is accumulated:

<u>PATIENTS</u>	<u>SIZE</u>	<u>COLOR</u>	<u>DECISIONS</u>
P1	.3L + .8S	.2R + .9B	.3/D _A + .6/D _B
P2	.4L + .7S	.4R + .7B	.8/D _A + .5/D _B
P3	.7L + .4S	.6R + .7B	.5/D _A + .9/D _B
P4	.8L + .5S	.3R + .8B	.7/D _A + .3/D _B
P5	.2L + .7S	.2R + .5B	.4/D _A + .2/D _B
P6	.9L + .2S	.8R + .2B	.7/D _A + .8/D _B
P7	.3L + .6S	.7R + .1B	.4/D _A + .5/D _B

with the following rules:

CERTAIN RULES:

If the tumor is large then D_A is present 0.5.

If the tumor is large and red then D_A is present 0.5.

If the tumor is large and blue then D_A is present 0.5.

If the tumor is red then D_B is present 0.5.

If the tumor is large and red then D_B is present 0.6.

If the tumor is small and red then D_B is present 0.5.

If the tumor is small and blue then D_B is present 0.5.

POSSIBLE RULES:

If the tumor is large then D_A is possible 0.7.

If the tumor is small then D_A is possible 0.7.

If the tumor is red then D_A is possible 0.7.

If the tumor is blue then D_A is possible 0.7.

If the tumor is large and red then D_A is possible 0.7.

If the tumor is large and blue then D_A is possible 0.7.

If the tumor is small and blue then D_A is possible 0.7.

If the tumor is large then D_B is possible 0.8.

If the tumor is small then D_B is possible 0.6.

If the tumor is red then D_B is possible 0.8.

If the tumor is blue then D_B is possible 0.7.

If the tumor is large and red then D_B is possible 0.8.

If the tumor is large and blue then D_B is possible 0.7.

If the tumor is small and blue then D_B is possible 0.6.

Finally, we extract the certain rules and possible rules in which to keep by using the theory explained in section 2.3. This leads to the following rules:

EXTRACTED CERTAIN RULES:

- If the tumor is large then D_A is present 0.5.
- If the tumor is red then D_B is present 0.5.
- If the tumor is large and red then D_B is present 0.6.
- If the tumor is small and blue then D_B is present 0.5.

EXTRACTED POSSIBLE RULES:

- If the tumor is large and red then D_A is possible 0.7.
- If the tumor is large and blue then D_A is possible 0.7.
- If the tumor is small and blue then D_A is possible 0.7.
- If the tumor is large and red then D_B is possible 0.8.
- If the tumor is large and blue then D_B is possible 0.7.
- If the tumor is small and blue then D_B is possible 0.6.

- SECTION 4 -

Software Design and Specifications

4.1 Software Specifications

The following is an attempt to describe the requirements specifications for the software to be developed for partial fulfillment of the senior project (CS 4395). The software should be designed to simulate the main ideas in Dr. Andre' de Korvin's paper, *"Extracting fuzzy rules under uncertainty and measuring definibility using rough sets."*

As in all good software design, the software should be above all user-friendly. It should be designed to allow a user to "walk-through" the system. This can be achieved through screen messages at every step and error messages when appropriate (improper data entry). The software should also be modifiable so that it may be expanded in the future. This can be achieved through well-documented modules. The software should also be efficient and reliable.

These are the goals of every software system. The following is a list of the functions, goals, and constraints of this particular system. In some instances, examples are used to better explain the concepts.

4.1.1 Input

The user should be able to:

A. Enter any number of attributes.

The paper uses two, for example: size and color. The user should also be allowed to have any number of descriptions for each attribute. The paper describes size with values of large and small. The user should also be able to use medium.

B. Enter data in any numeric form.

- 1) The form the data is entered in the paper is in "fuzzy form", where all values are between 0 and 1. The software should certainly be able to manipulate data which is entered in this form. In addition, the user should be able to enter "real data". For example, given the following numbers:

10	40	5	50
15	27	80	25
60	35	55	33

The software should be able to convert 55 to
55 -> .3/Low + .7/High

- 2) The user should also be able to set the boundaries for the data to be entered. Using the numbers from above, the user may wish to declare 10 as the bottom and 75 as the ceiling. If the number 5 is entered as data, it should be converted to: 5 -> 1/Low + 0/High. Likewise, 80 would be converted to: 80 -> 0/Low + 1/High. The user should be able to arbitrarily set these boundaries as well as change them between applications.

C. Set the two threshold values (one for the certain rules, one for the possible rules).

- 1) The user should be able to interactively change the threshold to compare the changes, i.e. the rules the changes affect.
- 2) Software should produce an error message for a threshold value greater than 1 or less than 0.

4.1.2 Functions and Calculations

The software should be able to:

- A. Convert the inputted data into the "fuzzy sets" that are used as the basis for all functions and calculations.
- B. Measure the degree to which a set, A, is included in another, B: $\{I(A \subset B)\}$. This calculation is used to determine the certain rules.
- C. Measure the degree to which a set, A, intersects another, B: $\{J(A \cap B)\}$. This calculation is used to determine the possible rules.
- D. Compare the values of $I(A \subset B)$, for various A's and B's, with the threshold value for certain rules and disregard all values of $I(A \subset B)$ which fall below the threshold. Similarly, all values of $J(A \cap B)$ should be compared to the threshold value for possible rules with all values of $J(A \cap B)$ below the threshold being disregarded.
- E. From the values of $I(A \subset B)$ and $J(A \cap B)$ that are at or above the threshold, the software should extract the rules (to keep). For the certain rules, the "prime" rules should be extracted. For the possible rules, the "combination" rules should be extracted. For example, if the rules are:
 - (1) If tumor is A and B then C is .6.
 - (2) If tumor is A then C is .6.
 - (3) If tumor is B then C is .6.For the certain rules, we extract (2) and (3). For the possible rules, we extract (1).
- F. Convert the inclusion $\{I(A \subset B)\}$ and intersection $\{J(A \cap B)\}$ symbols to english statements. The purpose of this is to help the user to better distinguish the output.

4.1.3 Output

The software should produce:

- A. A complete listing of all rules (certain and possible) in english for :
 - 1) Before comparison to the threshold value, and
 - 2) After the comparison to the threshold value.
- B. The two threshold values the user has assigned.
- C. The final list of extracted certain and possible rules in english.

References

- Chang, S.K., Wang, P.P. Fuzzy Sets: Theory and Application to Policy Analysis and Information Systems. Plenum Press, New York (1980).
- de Korvin, Andre', Bourgeois, B., Kleyle, R. *Extracting Fuzzy Rules under Uncertainty and Measuring Definability using Rough Sets*, Technical Report, University of Houston - Downtown, Houston, TX.
- de Korvin, A., Kleyle, R. *A Unified model for data acquisition and decision making*, Journal of the Amer. Soc. for Info. Science 15 pp. 149 - 161 (1989).
- de Korvin, A., Kleyle, R., Lea, R. *An evidential approach to problem solving when a large number of knowledge systems are available*, The Intern. Journal of Intelligent Systems 5, pp. 293 - 306 (1990).
- Dubois, D., Prade, H. Fuzzy Sets and Systems: Theory and Application. Radius Press, New York (1980).
- Ghezzi, C., Jazayeri, M., Mandrioli, D. Fundamentals of Software Engineering. Prentice Hall, New Jersey (1991).
- Grzymala-Busse, J.W. *Knowledge acquisition under uncertainty - a rough set approach*, Journal of Intelligent and Robotic Systems 1, pp. 3 - 16 (1988).
- Grzymala-Busse, J.W. *Dempster-Shafer theory interpretation of rough set approach to knowledge acquisition under uncertainty*, Tech. Rept., University of Kansas, Lawrence, KS (1991).
- Kachigan, S.K. Statistical Analysis: An interdisciplinary introduction to Univariate and Multivariate Methods. Radius Press, New York (1982).
- Mynatt, B.T. Software Engineering with Student Project Guidance. Prentice Hall, New York (1990).

- Pawlak, Z. *Rough Sets: Basic notation*, Institute Comp. Sci., Polish Acad. Sci. Rep No. 431, Warsaw (1981).
- Pawlak, Z. *Rough Sets*, Int. Journal Information Computer Sci. 11 pp. 341 - 356 (1982).
- Pawlak, Z. *Rough Sets and Fuzzy Sets*, Fuzzy sets and systems 17 pp. 99 - 102 (1985).
- Shafer, G. Mathematical Theory of Evidence. Princeton Press, New York (1976).
- Strat, T.M. *Decision analysis using belief functions*, International Journal of Approximate Reasoning 4, pp. 391 - 417 (1990).
- Yager, R.R. *Decision making under Dempster-Shafer uncertainties*, Tech. Report, Iona College. Report MII-915.
- Zadeh, L.A. *The rule of fuzzy logic in the management of uncertainty in expert systems*, Fuzzy sets and systems 11, pp. 119 - 227 (1983).

**EXTRACTING FUZZY RULES UNDER UNCERTAINTY
AND
MEASURING DEFINIBILITY USING ROUGH SETS**

**Prepared by
Donald E. Culas
in
Partial fulfillment
of the requirements
for
CS 4395**

**Department of Applied Mathematical Sciences
University of Houston - Downtown
November 22, 1991**

Committee Members / Approval

Dr. Andre' de Korvin	Faculty Advisor	-----
Dr. Kenneth Oberhoff	Member	-----
Dr. Robert Lea	Member	-----
Dr. Ananda Gunawardena	Member	-----
Dr. Richard A. Alo'	Chairman	-----

TABLE OF CONTENTS

Acknowledgements	i
Abstract	ii
Chapter 1. Introduction	1
1.1 The Problem at hand	2
1.2 Reasoning and Probability theory	3
1.2.1 Probability Theory	4
1.3 Why not Classical Probability Theory	5
1.4 Discussion	7
1.5 Scope	8
Chapter 2. Rough Sets	10
2.1 Basic Notations and Concepts	10
2.2 Information Systems	12
2.3 Rough Definability of a Set	13
2.4 An Example	14
Chapter 3. Fuzziness and Fuzzy sets	18
3.1 Fuzziness	18
3.1.1 IF ... Then rules	19
3.1.2 Symptoms	21
3.1.3 Uncertainty in Data	22
3.2 Fuzzy Sets	24
Chapter 4. Extracting Fuzzy rules and Measuring Definability	27
4.1 Functions on pairs of fuzzy sets	27
4.2 An Example	29
4.3 Extracting Rules	32
4.4 Measuring Definability	34
References	35

ACKNOWLEDGEMENTS

I would like to thank Dr. Andre' de Korvin for letting me work on an interesting project such as this and for giving me his time and attention. Furthermore let me take this opportunity to thank Dr. Kenneth Oberhoff for his valuable suggestions and comments given to me during the software design and development stages. Finally, I would like to thank Dr. Richard A. Alo' for giving me an opportunity to take part in such a project, and for all the constructive comments given to me throughout the semester.

Abstract

Although computers have come a long way since their invention, they are basically able to handle only crisp values. Unfortunately, the world we live in consists of problems which fail to fall into this category, i.e. uncertainty is all too common. In this work we look at a problem which involves uncertainty. To be specific we deal with attributes which are fuzzy sets. Under this condition we acquire knowledge by looking at examples. In each example a condition as well as a decision is made available. Based on the examples given to us, we will extract two sets of rules namely: certain and possible. Furthermore we will construct measures of how much we believe these rules, and finally we will define the decisions as a function of the terms used in the conditions.

CHAPTER 1

INTRODUCTION

Despite the advancements made in computer technology to date, all the computers on the market today are stored program sequential processing machines built around the Von Neumann architecture, the principles for which date back to the Turing machine, a computing model first proposed by Alan Turing in 1936. This means that, in principle, modern day computers are designed primarily to carry out mathematical calculations. Human expectations vis-a-vis computers know no bounds. These expectations go beyond routine jobs such as numerical calculations and the processing of office work, to include support in decision-making processes, the ability to understand natural languages, the diagnosis of malfunctions and the processing of intellectual information such as that required in design and planning work. To accomplish these kinds of operations, symbol processing computers equipped with inference functions are required. However, even symbol processing machines are not capable of handling experience and intuition, two very aspects of intelligence. This is because conventional computers are extremely **crisp**, (i.e. capable of dealing with definite values) having being designed around the binary logic of Boolean algebra. Human experience and intuition, however, by their very nature are multi-numerical. In other words, they are **fuzzy**.

This raises the question of just how necessary it is to have computers that are capable of processing ambiguous information, such as human experience and intuition. After all can't most phenomena encountered in this world be thoroughly processed mathematically? This question has been raised because today's computers are only used to solve well-structured problems for which all information is available. However, the everyday real world in which we live is rife with problems for which not

all information is available, and which are not well structured. This project focuses on such a problem, for which all information is not available and/or not well structured.

1.1 The Problem at Hand

Expert systems of a certain kind rely essentially upon the availability of a method for handling uncertainty. These systems cannot be conceived without a decision being firstly made about the choice of this method. Obviously this is true for all expert systems using empirical knowledge which in itself is not absolutely certain. As an example, we could mention a medical expert system which draws conclusions from the observed symptoms about whether or not a certain disease is present. All conclusions of this type inevitably contain an amount of uncertainty. However the rules which lead to these conclusions should not be confused with logical rules and must not be treated in the same way.

We shall call expert systems of this type **diagnostic systems**. They are mostly in the field of medicine, but can also be used in many other applications such as meteorology or geology, and of course for the control of technical installations. Therefore the expression **diagnostic system** should always be understood in the sense of an expert system, which relies upon empirical interdependencies for drawing its conclusions and consequently requires the treatment of uncertainty.

In order to make it possible to decide upon an appropriate therapy, a quantitative measure of uncertainty has to be applied in all relevant cases of a diagnostic system. Moreover it may be sensible to establish rules which, in certain stages of the investigation, direct the investigator's efforts depending on the degree of certainty achieved for possible hypothesis.

It is evident therefore, that for researchers who design diagnostic systems the question has to be answered, as to which method of measuring uncertainty should be employed. For more than three hundred years scientists, philosophers,

mathematicians and statisticians have used the concept of probability to describe degrees of uncertainty. Over three centuries a huge amount of theoretical results and experiences concerning the applicability of probability theory in different fields of human knowledge has been accumulated. Nevertheless many doubts concerning the appropriateness of the use of probability in diagnostic systems have arisen during the last decade. In the following sections we look at the probability theory and see why this is so.

1.2 Reasoning and Probability Theory.

Decision making often involves the use of rules. Simple rules are acceptable to most people in their everyday life, e.g.

In India, if you are under 60 years of age then you are entitled to a retirement position.

A rule for entitlement might be more complex, but understandable, e.g.

If you are at least 60 years old and female and you have been resident in India for at least 25 years, or if you are male and at least 65 years old and you have been resident in India for at least 30 years then, provided you are not receiving a disability pension, you are entitled to the retirement position.

People use 'if then....' statements in conversation, and often use rules in their everyday lives. However, problems which require expertise are not deterministic, i.e. the solutions cannot be stated in simple rules. Where judgement is involved, people often use words like probably, unlikely, almost certainly, i.e. uncertainty is involved. In some cases they quantify what they mean. For example:

I am 99% confident that if you water the plant, its condition will improve.

There is a small risk, about 5%, that you have this disease.

The ways in which people use these percentages are ill defined and often inconsistent. However, as we shall see in the next section, there is a mathematical theory of probability which provides a logical model for uncertainty.

1.2.1 Probability Theory.

Probability theory originated in the seventeenth century in the context of gambling. A gambler assesses his chance of winning and therefore the risk associated with his bid [3]. This process is very similar to that of an expert weighing up evidence, and judging whether he has sufficient evidence to justify a particular course of action. Chance, expectation and risk are components of both probability theory and expert judgements.

Probability is a measure of certainty between 0 and 1. The extreme values denote impossibility and certainty. Most people would understand that if a fair coin is tossed then the probability of its landing on a certain side is 0.5. This is because we ignore the possibility of its landing on its edge or not landing at all, and the other two outcomes are equally likely. Furthermore, only one of the events (head or tail) can occur at once, i.e. the events are **mutually exclusive**. This leads us to the classical definition of probability:

If a random experiment has N possible outcomes which are all equally likely and mutually exclusive, and n of these possibilities has outcome A then the probability of outcome A is n/N .

For example, consider a standard pack of 52 playing cards which has been shuffled so that the order of the cards is unpredictable. If a card is picked at random then the chance that it is a club is $13/52 = 0.25$. This is a very simplistic view of uncertainty. The definition depends on the terms random, mutually exclusive and equally likely. It cannot help much with questions like:

What is the probability that a child born in the United States will be a male?

What is the probability that the pain is caused by indigestion, and not a serious illness?

These are all real questions, and experts continually make similar judgements. If we looked at the record of births in the United States over the past two years then we could calculate the relative frequency of male births, i.e. the ratio of number of boys to number of births. We would expect this to be close to the true probability. Assuming that there had been no genetic changes, a more reliable estimate could be obtained from the records of the past ten years. So, if we can imagine a series of observations under constant conditions then the probability p of event A can be approximated by the relative frequency of A in a series of such observations. In practice 'true' probabilities are almost impossible to quantify, and most probabilities used are estimates based on relative frequencies.

1.3 Why not Classical Probability Theory?

Even though, the basic ideas prevailing in some considerations about diagnostic systems sound convincing, they violate fundamental requirements for reasonable handling of uncertainty. These ideas may be described as follows: If a certain fact is observed, a measure M_1 of uncertainty concerning the hypothesis in question must exist. If in addition another fact is observed, which produces a measure M_2 with respect to the same hypothesis, a combination rule must be given, which yields the measure of uncertainty of this hypothesis resulting from both observations. Such a rule, which calculates the measure of uncertainty for the combined observation as a function of the measures M_1 and M_2 can never take into account the kind of mutual dependence of the two observed facts. It might well be that these facts nearly always occur together, if indeed they occur at all. In such a situation the second observation is redundant and should not be used to update the measure of uncertainty. In another situation the two facts very seldom occur simultaneously and if they do, then this is an

important indication concerning the hypothesis in question. If they do occur simultaneously, the updating of the measure of uncertainty should have drastic consequences. Classical probability theory which treats these two situations equally cannot be considered useful.

Another argument against the probability theory is that : Is it justifiable to attribute a certain measure of uncertainty to the observation of a given fact, irrespective of the circumstances? For example let's take the example of a medical diagnostic system: If a symptom Z is observed, and a measure of uncertainty is used concerning the hypothesis of the presence of a certain disease, can this measure remain valid, if this disease occurs much more frequently than before? Once again an appropriate use of probability theory reveals the kind of dependence prevailing in this case. However, this will not be a popular result, because it states that a diagnostic system using this type of measure of uncertainty cannot be applied to populations showing different frequencies of this disease.

The problems of using probability models are compounded by the fact that people do not really understand the theory [6]. The theory itself is consistent and correct, but in order to apply it we need to make assumptions about underlying distributions and independence and sometimes use sophisticated mathematics to develop a consistent model for the system. Even given a consistent model people find it hard to estimate conditional probabilities. Statistical tests are a method of using probability theory to judge the weight of evidence and of selecting an hypothesis from two alternatives. However, many of the theorems and methods needed when using probabilities in diagnostic systems require the expert to estimate probabilities, sometimes without recourse to relative frequencies. Yet another problem with forcing experts to describe their inference in terms of probability theory is that the theory is not a natural method of reasoning.

1.4 Discussion

Recently, a lot of time and effort has been expended by the expert systems research community to the acquisition of knowledge under uncertainty. Uncertainty arises in many different situations. It may be caused by the ambiguity in the terms used to describe a specific situation, it might be caused by skepticism of rules used to describe a course of action or by missing and/or erroneous data.

In order to deal with uncertainty, techniques other than classical logic need to be developed. Statistics is the best tool available to handle likelihood. However, in many cases probabilities need to be estimated, sometimes without even recourse to relative frequencies. Estimates, then are likely to be very inaccurate. Many authors have cited theoretical weaknesses of expert systems based on statistical technique. In particular, there has been an attempt to create a system for the verification of indications for treatment of duodenal ulcers by HSV on the basis of statistics. The results were counter-intuitive and the system was rejected by physicians. The Dempster- Shafer theory of evidence or the theory of belief functions, give a useful measure for the evaluation of subjective certainty. The Dempster-Shafer theory has recently become popular. For an in depth-look at the Dempster-Shafer theory the reader is referred to [10]. Fuzzy logic, based on Zadeh's theory of **fuzzy sets** (where the degree to which an optional element (a) belongs to set (A) is determined by assigning it a value or grade ranging from 0 to 1) is another means of handling uncertainty. However, this too has problems [9]. There is extensive literature on ways to deal with uncertainty in expert systems, like a combination of statistics and fuzzy logic, theory of endorsements [1], nonmonotonic logic [7, 8], modal logic etc. [5].

One of the most popular ways to acquire knowledge is based on learning from examples. An effective tool to infer knowledge from examples is **rough set theory**. Rough set theory was introduced in 1981 by Z. Pawlak as a method to acquire knowledge under uncertainty. The main assumption of the rough set theory is that the

information stored in the data base like system, called an information system, may contain inconsistencies. In the process of acquiring knowledge these inconsistencies are taken into account. Thus, using the basic tools of rough set theory, which we will look at closely in the next chapter, two sets of rules are produced namely **certain** and **possible**. The main advantage of the rough set theory is that it does not need any preliminary or additional information about data like probability in statistics. Moreover rough set theory has been successfully implemented in knowledge-based systems in medicine and industry. In particular, an expert system based on rough set theory for engineering design is being developed at Wayne State University, Michigan, and University of Regina, Canada.

1.5 Scope.

In this work, we will deal with a setting where a decision maker is faced with uncertain (i.e. fuzzy) symptoms and makes a fuzzy diagnosis which might be strongly or weakly based on these symptoms. The cases which we will look at are not "textbook cases" and the values of attributes are not crisp. Moreover the diagnosis is not of a "pure type". It is a mixture of several "pure types". Thus, a patient might have a diagnosis of the type $.3/D_A + .6/D_B$ meaning that the physician believes the fuzzy symptoms reflect disease D_A with strength .3 and disease D_B with strength .6. From such a setting we will extract fuzzy rules using the rough set theory.

Fuzzy rules are naturally present in descriptions, crisp rules are the exceptions. Also, fewer fuzzy rules are needed than crisp ones to build an expert system. Thus a rule such as : If the tumor is somewhat large then the presence of skin cancer is somewhat likely is the type of rule experts naturally use as opposed to giving the size of a tumor and a number expressing the probability of cancer.

In the first part of this work we will develop a methodology to extract rules such as the ones stated above, from fuzzy symptoms and fuzzy diagnosis. In fact we will

extract two sets of rules i.e. **certain** and **possible** rules as well as a measure of how much we believe these rules. In the second part we will look at a related problem that is to define the diagnosis in terms of the symptoms. In the next chapter we take an in-depth look at the rough set theory which is necessary to understand the rest of this paper.

CHAPTER 2

ROUGH SETS

Acquiring knowledge under uncertainty is one of the main problems of expert systems. One of the most popular ways to acquire knowledge is based on learning from examples. In 1981, Z. Pawlak introduced a new tool, namely rough set theory to acquire knowledge under uncertainty. In this chapter we look at the basic concepts of rough set theory. Other methods have been developed prior to the introduction of rough set theory. However, use of the rough set theory seem to have many advantages over the other methods. One of the main advantages of the rough set theory is that it does not need any preliminary or additional information about data (like probability in statistics, basic probability number in Dempster-Shafer theory, grade of membership, or the value of possibility in fuzzy set theory). Another advantage of rough set theory is that its algorithms are very simple, and the theory itself is clear and easy to follow. Moreover the theory has been successfully implemented in many cases in expert systems in medicine and industry.

2.1 Basic Notations and Concepts

All the concepts mentioned in this section can be found in [4]. Let U be a nonempty set, called the **universe**, and let R be an equivalence relation on U called an **indiscernibility** relation. An ordered pair $A = (U, R)$ is called an **approximation space**. For an element x of U , the equivalence class of R containing x will be denoted by $[x]_R$. Equivalence classes of R are called **elementary** sets in A . We assume that the empty set is also elementary. Any finite union of elementary sets in A is called a **definable** set in A .

Let X be a subset of U and we wish to define X in terms of definable sets in A . Thus, we need two more concepts. A lower approximation of X in A , denoted by $\underline{R}X$, is the set given by

$$\{x \in U \mid [x]_R \subseteq X\}.$$

An upper approximation of X in A , denoted by $\overline{R}X$, is the set given by

$$\{x \in U \mid [x]_R \cap X \neq \emptyset\}$$

The lower approximation of X in A is the greatest definable set in A , contained in X . The upper approximation of X is the least definable set in A containing X . A rough set in A (or rough set, if A is known) is the family of all subsets of U having the same lower and upper approximations in A .

Let X and Y be subsets of U . Lower and upper approximations of X and Y in A have the following properties:

$$\underline{R}X \subseteq X \subseteq \overline{R}X,$$

$$\underline{R}U = U = \overline{R}U,$$

$$\underline{R}\emptyset = \emptyset = \overline{R}\emptyset,$$

$$\underline{R}(X \cup Y) \supseteq \underline{R}X \cup \underline{R}Y,$$

$$\overline{R}(X \cup Y) = \overline{R}X \cup \overline{R}Y,$$

$$\underline{R}(X \cap Y) = \underline{R}X \cap \underline{R}Y,$$

$$\overline{R}(X \cap Y) \subseteq \overline{R}X \cap \overline{R}Y,$$

$$\underline{R}(X - Y) \subseteq \underline{R}X - \underline{R}Y,$$

$$\overline{R}(X - Y) \supseteq \overline{R}X - \overline{R}Y,$$

$$\underline{R}(-X) = -\overline{R}X,$$

$$\overline{R}(-X) = -\underline{R}X,$$

$$\underline{R}X \cup \overline{R}(-X) = X,$$

$$\underline{R}(\underline{R}X) = \overline{R}(\underline{R}X) = \underline{R}X,$$

$$\overline{R}(\overline{R}X) = \underline{R}(\overline{R}X) = \overline{R}X,$$

where $-X$ denotes the complement $U-X$ of X .

Let x be an element of U . We define two additional membership relations $\underline{\epsilon}$ and $\overline{\epsilon}$, called **strong** and **weak memberships**, in the following way

$$x \underline{\epsilon} X \text{ iff } x \in \underline{R}X$$

and

$$x \overline{\epsilon} X \text{ iff } x \in \overline{R}X$$

with meanings: x is **certainly** in X and x is **possibly** in X respectively. Our terminology originates in that we want to decide if x is in X on the basis of definable sets in A rather than on the basis of X . This means that we deal with $\underline{R}X$ and $\overline{R}X$ instead of X , and since $\underline{R}X \subseteq X \subseteq \overline{R}X$, if x is in $\underline{R}X$ it is certainly in X . On the other hand, if x is in $\overline{R}X$, it is possibly in X .

2.2 Information Systems

An information system is similar to a data base. The difference is that the entities of such an information system, called objects, do not need to be distinguished by attributes. The information system serves as the basis for knowledge acquisition, producing rules from examples. Therefore, attributes are divided into two types: conditions and decisions (or actions). Objects are described by values of conditions, while classifications made by experts are represented by values of decisions.

For example, if the system is a hospital, the objects would be patients, the condition attributes would be tests, and the decision attributes would be diseases. Each patient would be characterized by test results and would be classified by physicians (experts) as being on some level of disease severity. As another example if the system is an industrial process, the objects would be sample of processes taken at some specific moments in time. Conditions would be the parameters of the process, while the decisions would be actions taken by the operators (experts).

An information system S is a quadruple (U, Q, V, P) where U is a nonempty finite set, and its elements are called objects of S , $Q = C \cup D$ is a set of attributes, C is a

nonempty finite set, its elements are called condition attributes of S , and D is also a nonempty finite set, and its elements are called decision attributes of S , $D \cap C = \emptyset$, $V = \bigcup_{q \in Q} V_q$ is a nonempty finite set, and its elements are called values of attributes, where V_q is the set of values of attribute q , called the domain of q , and p is a function of $U \times Q$ onto V , called a description of S , such that $p(x, q) \in V_q$ for all $x \in U$ and $q \in Q$.

Let P be a nonempty subset of Q , and let x, y be members of U . Objects x and y are indiscernible by P in S , denoted by $x \tilde{P} y$, iff for each q in P , $p(x, q) = p(y, q)$. Obviously, \tilde{P} is an equivalence relation on U . Thus P defines a partition on U ; such a partition is a set of all equivalence classes of \tilde{P} . This partition is called a classification of U generated by P in S , or briefly a classification generated by P .

2.3 Rough Definability of a Set

For a nonempty subset P of Q , an ordered pair (U, \tilde{P}) is an approximation space A . For the sake of convenience, for any $X \subseteq U$, the lower approximation of X in A and the upper approximation of X in A will be called P -lower approximation of X in S and P -upper approximation of X in S , and will be denoted by $\underline{P}X$ and $\overline{P}X$, respectively. A definable set X in A will be also called P -definable in S . Thus, X is P -definable in S iff $\underline{P}X = \overline{P}X$.

For a nonempty subset P of Q , a set $X \subseteq U$ which is not P -definable in $S = (U, Q, V, P)$ will be called P -undefinable in S . Set X is P -undefinable iff $\underline{P}X \neq \overline{P}X$.

The set X will be called roughly P -definable in S iff $\underline{P}X \neq \emptyset$ and $\overline{P}X \neq U$.

The set X will be called internally P -undefinable in S iff $\underline{P}X = \emptyset$ and $\overline{P}X \neq U$.

The set X will be called externally P -undefinable in S iff $\underline{P}X \neq \emptyset$ and $\overline{P}X = U$.

The set X will be called totally P -undefinable in S iff $\underline{P}X = \emptyset$ and $\overline{P}X = U$.

For an internally P -undefinable set X in S we can not say with certainty that any $x \in U$ is a member of X . For an externally P -undefinable set X in S we cannot

exclude any $x \in U$ being possibly a member of X . In the next section we look at an example which illustrates the above mentioned concepts.

2.4 An Example

Let us look at the information system which is given by the following table.

Table 1. An example of information system:

	C		
	C		D
	Temperature	Headache	Influenza
X1	normal	no	no
X2	normal	yes	no
X3	normal	yes	yes
X4	subfebrile	no	no
X5	subfebrile	yes	no
X6	subfebrile	yes	yes
X7	high	no	yes
X8	high	yes	yes
X9	high	yes	yes

The classification, generated by the set C of conditions attributes, called **Temperature** and **Headache**, is equal to

$$\{ \{X_1\}, \{X_2, X_3\}, \{X_4\}, \{X_5, X_6\}, \{X_7\}, \{X_8, X_9\} \}.$$

The set D of decision attributes consists of one member, called **Influenza**.

As can be seen in the table, an expert introduced two inconsistencies. First, he assigned different values of condition attributes to patients x_2 and x_3 , in spite of the fact that both patients, x_2 and x_3 , characterized by the same values of condition attributes **Temperature** and **Headache**. Yet another inconsistency is associated with patients x_5 and x_6 .

Let us assume that $X = \{x \mid p(x, d) = \text{no}\}$, i.e. $X = \{x_1, x_2, x_4, x_5\}$. Thus X represents all patients in U , classified by an expert in the same way, as being not sick with influenza. Then

$$\underline{C}X = \{x_1\} \cup \{x_4\} = \{x_1, x_4\},$$

$$\overline{C}X = \{x_1\} \cup \{x_2, x_3\} \cup \{x_4\} \cup \{x_5, x_6\} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

It is the presence of inconsistencies that produce a difference between the lower and upper approximation.

In our example, $\underline{C}X \neq \emptyset$ and $\overline{C}X \neq U$, therefore X is roughly C-definable in S . For set X , sets $\underline{C}X$ and $\overline{C}X$ are illustrated by the following figures:

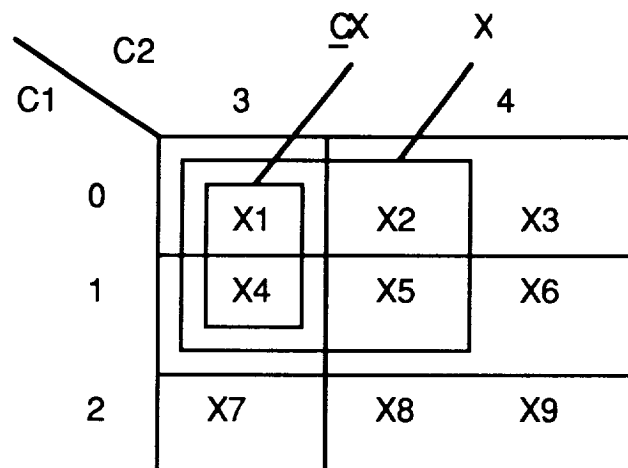


Figure (a) lower approximation $\underline{C}X$ of set X

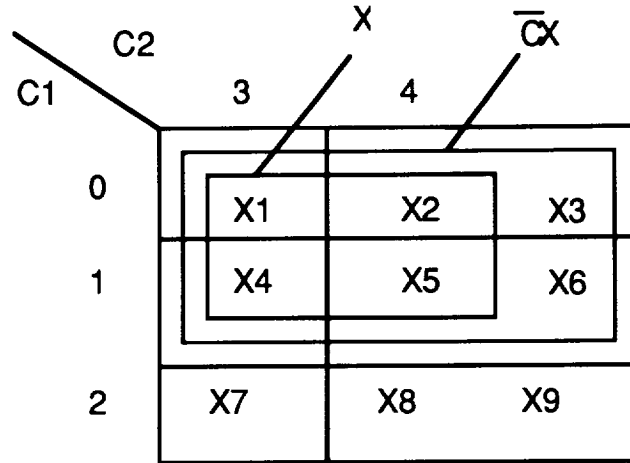


Figure (b) upper approximation \overline{CX} of set X .

The set X determines the following rough set:

{ $\{x_1, x_2, x_4, x_5\}$, $\{x_1, x_2, x_4, x_6\}$, $\{x_1, x_3, x_4, x_5\}$, $\{x_1, x_3, x_4, x_6\}$ }.

For example, $x_1 \in X$, hence $x_1 \in \overline{X}$, and $x_3 \notin X$, but $x_3 \in \overline{X}$.

Now let us represent the decision of the expert from the example, corresponding to set X , by rules. Any such a rule is a conditional statement that specifies a decision under conditions. The smallest subsets of U which may be described by rules, using the set C of conditions, are the members of the classification generated by C . Therefore, we may represent set X by rules iff X is C -definable. If set X is C -undefinable we cannot represent it by a single set of rules. Instead, we may represent sets \underline{CX} and \overline{CX} by different sets of rules. In particular a rule derived from \underline{CX} is **certain**, and a rule derived from \overline{CX} is **possible**.

In the example, X is roughly C -definable in S . The **certain** rules, corresponding to set \underline{CX} of positive examples and set \overline{CX} of negative examples, are

(Temperature, low) \rightarrow (Influenza, no)

(Temperature, subfebrile) \wedge (Headache, no) \rightarrow (Influenza, no),

and the **possible** rules, corresponding to set \overline{CX} of positive examples and set $\overline{\overline{CX}}$ of negative examples, are

(Temperature, low) --> (Influenza, no)

(Headache, no) --> (Influenza, no).

As can be seen from the above example, uncertainty is all too often present in the conditions and the decisions. The conditions and the decisions fail to partition the universe into well defined classes and some overlap is present. In real cases we do not have sharp boundaries between say **normal**, **subfebrile**, and **high**. The best we can hope is that **normal**, **subfebrile**, and **high**, "somewhat partition" the universe by not overlapping "too much." In the next chapter we will look at a method which would help us deal with such a setting, where attributes fail to have sharp boundaries.

CHAPTER 3

FUZZINESS AND FUZZY SETS

The way we humans actually store and manipulate concepts in the mind is a subject of some debate. However, we communicate conscious processes to other people verbally. This type of reasoning is done with words rather than numbers. This is why we find probability theory counter-intuitive, and in some cases difficult to understand. The many shades of meaning which give language its richness and colour contrast with the precise rigour of mathematical theory, logic and computer languages.

There is a difference between the meaning and usage of words. It is not the strict dictionary definition of a word which is important, but the way in which an expert uses a word. At times an expert may find it difficult to define a particular word, though usually he will be able to give an example of a use. We usually find technical terms relatively easy to define. However, commonly used words are less easy to define, either in abstract or even in context. For example let's consider the word "cold". What is the criterion for saying that the weather is cold? The answer depends on factors like temperature and the time of year. For instance a cold summer's day can be milder than a warm winter's day. It is relatively easy to quote examples of cold days and days which are not cold. There is a vagueness or **fuzziness** about a certain range of temperatures; they might constitute coldness, and they might not. In this chapter we take an in-depth look at this aspect of vagueness or fuzziness.

3.1 Fuzziness

Everyone uses fuzzy words in their everyday lives, and seldom question whether they or others understand their usage of those words. An individual may not

be consistent in his own use of words, and there is even less chance that someone else has the same usage. Nevertheless, we all use words expressing belief when we are reasoning or arguing. For example, let us look at a quote from a doctor:

"I wouldn't expect that disease in a young girl of 20. It's so rare as to be negligible. It isn't worth carrying out the test on a young person. If they're young I'd most likely not do the tests. If they are older I probably would do them."

Here the doctor is using vague rules. Some fuzzy words which he uses are young, older, negligible, so rare, most likely, probably etc. When pressed to define such words, experts often find it extremely difficult.

There is also a distinction between uncertainty and imprecision, which is not always reflected in the models used in computer systems. Uncertainty refers to something which is not known for sure, and imprecision refers to something whose value is not known accurately. Statements can be uncertain, imprecise or both. For example:

"There will definitely be a rise in temperature: somewhere between 10 degrees and 25 degrees."

is imprecise but certain whereas:

"I think you should leave it on. If so you should set it to 180 degrees."

is precise but uncertain.

3.1.1 IF THEN rules

Crisp mathematical rules can be easily defined. The basis for a rule is:

IF A then B or $A \rightarrow B$

This states that if A is true, then B is necessarily true too. It does not state that B implies A, and B can be true with A false. It is difficult to find a clear example of this concept except in the context of mathematics, for example:

if $X = 2$ then $X^2 = 4$

Note that $X^2 = 4$ does not mean that X is the value 2; $X = -2$ is another solution. The rule is exactly equivalent to:

Not B \rightarrow Not A

Unfortunately, in common usage "if" and "only if" are interchanged and used improperly all too often.

Statements based on logic are made more complex by the use of **AND** and **OR**. **AND** is easy to understand, but **OR** is ambiguous. If a child is told "You can have sweets or an ice cream", the child will usually understand that she is not allowed both. This is an exclusive **OR**. The statement "The leaves on the tree are green or yellow" implies that possibly some leaves are green and others are yellow. This is an inclusive **OR**: Yellow and green can occur together. The English language does not distinguish between these two meanings, and the interpretation may depend on the context. In formal logic and computer logic, the inclusive **OR** is more common. Further ambiguities arise when both terms **AND** and **OR** are used in the same statement. For example let us consider the rule:

"If the patient is over 40 and has high blood pressure or is female then I would refer them."

Does this statement mean:

"If the patient is over 40 and has high blood pressure or if the patient is over 40 and is female then I would refer them."

or does it mean:

"If the patient is over 40 and has high blood pressure or if the patient is female then I would refer them."

Only the person who made the statement can identify the correct interpretation. Note that the two interpretations give potentially different outcomes for a female patient under the age of 40. Again, in computer logic the meaning is unambiguous - the problem arises because of the way we use words.

3.1.2 Symptoms

Much judgement and reasoning using vague rules involve weighing up the strength of evidence in symptoms. For example:

"Meniere's disease causes spells of dizziness."

is a rule of the form:

If A then B

i.e. if you have Meniere's disease then you will have spells of dizziness. If we are told that a patient has spells of dizziness then it is more credible that he has Meniere's disease. However, dizziness can be caused by other illnesses or disorders. If dizziness is a common ailment for this type of patient then we do not have much evidence for Meniere's disease, but if it is rare except as a consequence of the disease, then our inference is stronger. The strength of our inference depends on how likely B is in itself. If B is very common, then we have little evidence for A; if B is very rare then A becomes much more credible. So B is true makes A more credible is our vague rule.

In practice, there is usually more than one symptom, or evidence, i.e. the rule is :

$A \rightarrow B_1, B_2, \dots, B_n$

For example:

"Meniere's disease causes spells of dizziness, tinnitus, and progressive hearing loss."

This form of reasoning is the one which is often represented by **Baye's rule**. The weights of evidence used in the doctor's diagnosis are not independent; it is a combination of symptoms which gives credibility to the solution.

3.1.3 Uncertainty in Data

The vagueness or uncertainty which is an intrinsic feature of judgement is not unique to rules. Data presented to an expert or expert system can also be uncertain. Some data are clear facts with a yes/no answer, for example:

The applicant is over the age of 18

but others may be fuzzy:

The patient may have suffered from indigestion

So expertise involves dealing with uncertain data, and uncertain inference rules using that data. Much of the skill in judgement lies in weighing up the relative merits of data, facts guesses and hypothesis, etc., and using a plausible line of reasoning with them. There are essentially two aspects to this uncertainty: belief and value. Belief is analogous to probability and measures the level of credibility whereas probability is a numerical measure. People generally use words to express belief. There are over 50 terms in the English language expressing belief, and the number can be increased by qualifiers such as very, extremely etc. However, if a subset of these terms could be agreed upon, together with an hierarchy expressing the relationships between them, then there is no reason why the expert should not be able to express his knowledge in simple English which is natural to him. For example the figure on the next page shows a simple hierarchy showing the relationships between terms such as possible, certain and definite. A term low down on the hierarchy is stronger than the one higher up. So 'certain' is stronger than 'probable', and proved implies 'definite'. The main problem with this is ascertaining whether the expert is consistent in his usage of words, and whether the agreed relationships make sense to other people. The other element, that of value, is analogous to risk. Terms expressing value are those such as fatal, serious, dangerous, undesirable, etc. A possibility which is considered likely and serious may warrant immediate investigation, whereas one which is highly probable and undesirable may not. It will be necessary to draw up similar diagrams representing

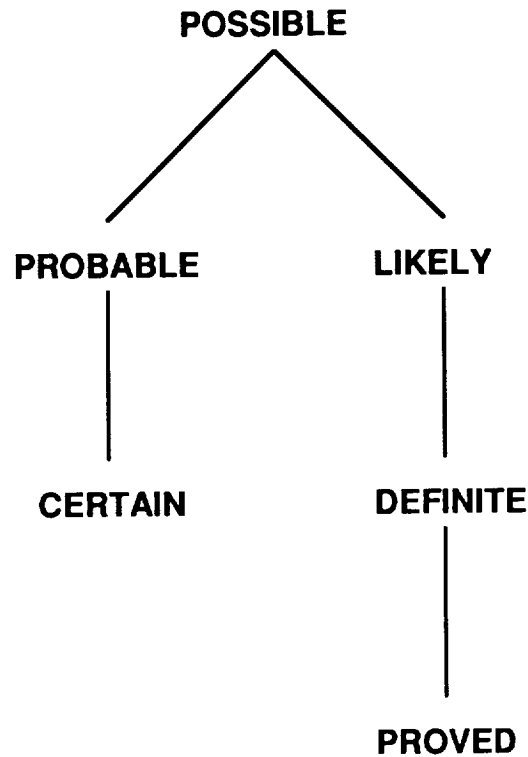


Figure (a) Hierarchy chart representing relationships between words.

relationships between words describing risk or value as well, if the uncertainty handling is to be written in words. If the expert can do this then it will usually be a valuable exercise. The problem, with using words is that there is such an abundance to choose from, but the advantage is that the language is easy for the expert to use. Risk is extremely important in reasoning processes. A low probability high risk situation might warrant investigation before a high probability low risk one. It is the importance which matters. Reasoning seems to be multi-dimensional and probability theory on its own seldom provides an adequate framework. In other words objective probabilities do not embrace all facets of human judgement.

3.2 Fuzzy Sets

Even though numerical models for belief have many disadvantages, it cannot be denied that many famous expert systems do use them. Pure probability theory has been considered inadequate and some famous systems use certainty factors. Another important theory which is used in expert systems is the **fuzzy set theory**. Fuzzy set theory and fuzzy logic were formulated by Zadeh, and have since been applied to many problems where traditional crisp logic and mathematics are inappropriate because of the inherent uncertainty. In traditional logic a proposition is true or false; in fuzzy logic it has a degree of truth. For example, let us consider the question:

Is the object black? (or white)

In crisp logic the answer can be either yes (black) or no (white). In fuzzy logic an object would be given a degree of blackness, where 0 indicates 'definitely not' and 1 indicates 'definitely'. An off-white object could be measured by 0.2, say, and a grey object by 0.6. This would not mean that one was three times as black as the other, but would enable the members of a set to be ranked. Let U be the universe of discourse or domain:

$$U = U_1 + U_2 + \dots + U_n$$

So U is the set of n objects U_1, U_2, \dots, U_n which we are considering. A fuzzy set F is described by its members and their degrees of membership to that set, for example:

$$F = M_1/U_1 + M_2/U_2 + \dots + M_n/U_n$$

U_1, U_2, \dots, U_n are members with degrees of membership M_1, M_2, \dots, M_n , and $+$ denotes union not addition. In other words this equation is a way of listing the various members together with their degrees of membership. Equivalently, F is given by:

$$F = \Sigma M_F(U_i)/U_i$$

where Σ denotes 'the set of'. We also define the fuzzy versions of union (inclusive OR), intersection(AND) and complement (NOT).

The grade of membership of U in the union $F \cup G$ (F OR G) is at least that of its membership in the individual sets F, G. We do not know any more than this, and so the grade of membership is given by the maximum of the two. So:

$$F \cup G = \Sigma M_F(U) \vee M_G(U) / U$$

where \vee denotes maximum. The grade of membership of U in $F \cap G$ (F AND G) can be no greater than the membership in each of F and G. So intersection is defined by:

$$F \cap G = \Sigma M_F(U) \wedge M_G(U) / U$$

where \wedge denotes minimum. The value 1 denotes full membership and 0 no membership.

The complement of F, F' is given by :

$$F' = \Sigma (1 - M_F(U)) / U.$$

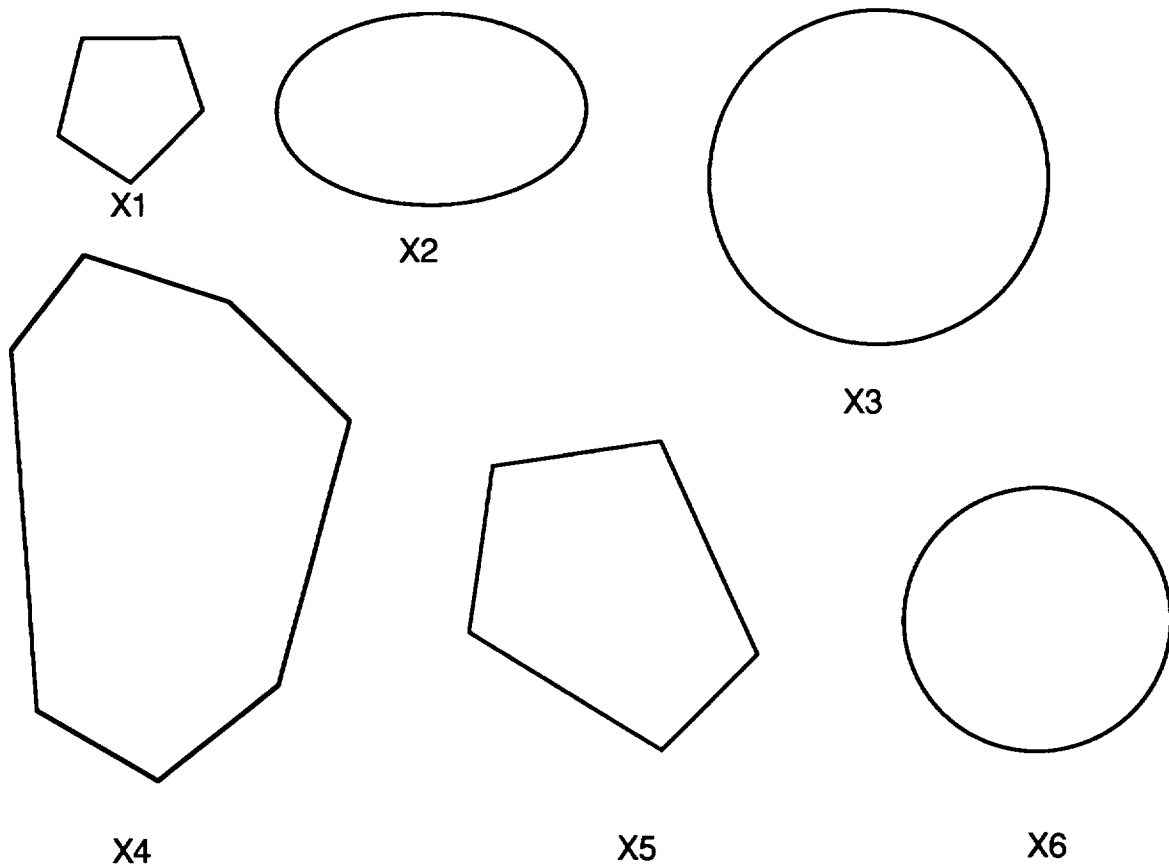


Figure (b) Set of six figures

Now let us look at an example by considering the objects in figure(b). Suppose L is the fuzzy set of large shapes, and R the fuzzy set of round shapes then L and R could be defined by:

$$L = 0.1/X_1 + 0.6/X_2 + 0.6/X_3 + 0.8/X_4 + 0.4/X_5 + 0.2/X_6$$

$$R = 0.1/X_1 + 0.7/X_2 + 1.0/X_3 + 0.5/X_4 + 0.1/X_5 + 1.0/X_6$$

$L \cup R$ is the set of objects which are large or round. X_1 is not really large and not particularly round, so its membership in $L \cup R$ is low. X_6 is not large but perfectly round so its membership in large or round is 1.

$$L \cup R = 0.1/X_1 + 0.7/X_2 + 1.0/X_3 + 0.8/X_4 + 0.4/X_5 + 1.0/X_6$$

$L \cap R$ is the set of objects which are large and round

$$L \cap R = 0.1/X_1 + 0.6/X_2 + 0.6/X_3 + 0.5/X_4 + 0.1/X_5 + 0.2/X_6$$

In this case X_5 and X_6 have low membership values for large and round ($L \cap R$) because membership in at least one of L and R is low. The strongest membership is for X_2 and X_3 both of which have fairly high membership in both L and in R together.

L' is the set of not large (i.e. small) objects

$$L' = 0.9/X_1 + 0.4/X_2 + 0.4/X_3 + 0.2/X_4 + 0.6/X_5 + 0.8/X_6$$

So X_1 has a high membership in L' and X_4 has a low membership. The different membership values mean that it is not sensible to count the members in a fuzzy set.

Instead we can define the power P of a set F by:

$$P(F) = \sum M_F(X)$$

So in our figure(b):

$$P(L) = 2.7$$

$$P(R) = 3.4$$

$$P(L \cup R) = 4.0$$

In the next chapter we discuss the concepts of the present work, which is to extract fuzzy rules under uncertainty and to measure definability using rough sets.

CHAPTER 4

EXTRACTING FUZZY RULES AND MEASURING DEFINIBILITY

The main purpose of the present work is to study a setting where a decision maker (expert) is faced with uncertain (i.e. fuzzy) symptoms and makes a fuzzy decision. Let's keep in mind that these decisions may be strongly or weakly based on the conditions. From the data we have, we will extract fuzzy rules, in fact we will extract two sets of rules i.e. **certain** and **possible** rules as well as a measure of how much we believe these rules. Finally we will define the decisions in terms of the symptoms. Before we go any further we will look at the properties, notations and operations of fuzzy sets which is required to understand this work.

4.1 Functions on pairs of fuzzy sets

We now look at some functions and properties of fuzzy sets. All the concepts explained here can be found in [2]. Let's recall from the previous chapter that a fuzzy subset A of U is defined by a characteristic function

$$\mu_A : U \longrightarrow [0,1].$$

The notation

$$\sum_i \alpha_i / x_i \quad (0 \leq \alpha_i \leq 1)$$

denotes a fuzzy subset whose characteristic function at x_i is α_i .

Moreover let us recall that if A and B are fuzzy subsets $A \cap B$, $A \cup B$, $\neg A$ are defined by $\min\{\mu_A(x), \mu_B(x)\}$, $\max\{\mu_A(x), \mu_B(x)\}$, and $1 - \mu_A(x)$ respectively. The implication $A \longrightarrow B$ is defined by $\neg A \cup B$ and the characteristic function corresponding to $\neg A \cup B$ is given by

$$\text{Max}\{ 1 - A(x), B(x) \}.$$

Let us now go through an example and see how these work:

$$\text{Let } A = (.5, .7, .2, .4) \text{ and}$$

$$\text{Let } B = (.4, .8, .9, .6)$$

then we have the following :

$$\text{Min}\{ \mu_A(x), \mu_B(x) \} = (.4, .7, .2, .4)$$

$$\text{Max}\{ \mu_A(x), \mu_B(x) \} = (.5, .8, .9, .6)$$

$$1 - \mu_A(x) = (.5, .3, .8, .6)$$

Now we look at two new functions on pairs of fuzzy sets.

$$I(A \subset B) = \inf_x \text{Max}\{ 1 - A(x), B(x) \}$$

$$J(A \# B) = \text{Max}_x \text{Min}\{ A(x), B(x) \}.$$

where A and B denote fuzzy subsets of the same universe. The function $I(A \subset B)$ measures the degree to which A is included in B and the function $J(A \# B)$ measures the degree to which A intersects B. If A and B are crisp sets it is evident that

$$I(A \subset B) = 1 \text{ if and only if } A \subset B$$

otherwise it is 0.

Moreover in the case of crisp sets

$$J(A \# B) = 1 \text{ if and only if } A \cap B \neq \emptyset$$

otherwise it is 0.

In addition to the above, lets also look at the following relation as shown in [2]

$$I(A \subset B) = 1 - J(A \# \neg B).$$

The right hand side of the above equation is

$$\begin{aligned} & \inf_x \{ 1 - \text{Min}\{ A(x), 1 - B(x) \} \} \\ &= \inf_x \text{Max}\{ 1 - A(x), 1 - (1 - B(x)) \} \end{aligned}$$

$$= \inf_x \text{Max} \{1 - A(x), B(x)\}.$$

In the next section we go through an example and show step by step how fuzzy certain and possible rules can be extracted from raw data.

4.2 An Example

Let us consider the following table which is the kind of raw data we will be dealing with in this work, i.e. we will have a set of conditions and a set of decisions whose values are given using the fuzzy set theory.

Table 1

<u>PATIENTS</u>	<u>SIZE</u>	<u>COLOR</u>	<u>DECISIONS</u>
P1	3/L + .8/S	.2/R + .9/B	.3/D _A + .6D _B
P2	.4/L + .7/S	.4/R + .7/B	.8/D _A + .5/D _B
P3	7/L + .4/S	.6/R + .7/B	.5/D _A + .9/D _B
P4	.8/L + .5/S	.3/R + .8/B	.7/D _A + .3/D _B
P5	.2/L + .7/S	.2/R + .5/R	.4/D _A + .2/D _B
P6	.9/L + .2/S	.8/R + .2/B	.7/D _A + .8/D _B
P7	.3/L + .6/S	.7/R + .1/B	.4/D _A + .5/D _B

L = Large

R = Red

D_a = Disease A

S = Small

B = Blue

D_b = Disease B

We will interpret the above table as a case where an expert is trying to determine the presence or absence of a disease by looking at the size and color of a tumor. The first

column represents a number of patients i.e. P₁, P₂,...,P₇. The symbols L and S stand for large and small respectively, and the symbols R and B stand for Red and Blue respectively. So we can interpret that patient P₁ has a tumor that is judged to be .3 large and .8 small. In this particular case, let us assume that a number of physicians are looking at the tumor, and that a certain number of them judge the tumor to be large, and others judge it to be small. So in our case the numbers .3 and .8 denote relative frequencies. However, it does not need to be so, i.e. these numbers could reflect some judgement and need not be generated as relative frequencies. The decision column shows a fuzzy diagnosis. So from our table, one's interpretation could be that patient P₁ is diagnosed to have disease D_A and the corresponding belief is .3. Also patient P₁ is diagnosed to have disease D_B and the corresponding belief is .6 strong.

Now what we want to do is to take these cases and unravel them into fuzzy rules as to when disease D_A or D_B is present. The first step is to take this raw data and convert them into fuzzy sets as follows:

$$D_A = .3/X_1 + .8/X_2 + .5/X_3 + .7/X_4 + .4/X_5 + .7/X_6 + .4/X_7$$

The fuzzy set for D_A is obtained by taking the union of the values of D_A of all the patients. Similarly fuzzy sets are created for large, small, red and blue as follows:

$$L = .3/X_1 + .4/X_2 + .7/X_3 + .8/X_4 + .2/X_5 + .9/X_6 + .3/X_7$$

$$S = .8/X_1 + .7/X_2 + .4/X_3 + .5/X_4 + .7/X_5 + .2/X_6 + .6/X_7$$

$$R = .2/X_1 + .4/X_2 + .6/X_3 + .3/X_4 + .2/X_5 + .8/X_6 + .7/X_7$$

$$B = .9/X_1 + .7/X_2 + .7/X_3 + .8/X_4 + .5/X_5 + .2/X_6 + .1/X_7$$

The next step would be to find the minimum degree to which possible combinations of symptoms imply disease D_A i.e. find the certain rules. This is done by computing $I(L \subset D_A)$, $I(S \subset D_A)$, $I(R \subset D_A)$, $I(B \subset D_A)$, $I(L \cap R \subset D_A)$, $I(L \cap B \subset D_A)$, $I(S \cap R \subset D_A)$. Similar computations would be carried out for D_B.

Carrying out the computations would yield the following results:

$$I(L \subset D_A) = .5 \quad I(L \subset D_B) = .3 \quad I(S \subset D_A) = .3$$

$$\begin{array}{lll}
I(S \subset D_B) = .3 & I(R \subset D_A) = .4 & I(R \subset D_B) = .5 \\
I(B \subset D_A) = .3 & I(B \subset D_B) = .3 & I(L \cap R \subset D_A) = .5 \\
I(L \cap R \subset D_B) = .6 & I(L \cap B \subset D_A) = .5 & I(L \cap B \subset D_B) = .3 \\
I(S \cap R \subset D_A) = .4 & I(S \cap R \subset D_B) = .5 & I(S \cap B \subset D_A) = .3 \\
I(S \cap B \subset D_B) = .5 & &
\end{array}$$

Now all these yield **certain** rules. But we may not want to keep all the rules in order to avoid any partial implications. So we would set a threshold value, say for this example let us choose threshold value (α) to be .5. This would throw away any rule which evaluates below this threshold. Of course the lower the α is, the more partial implications are taken into account. The choice of α is very much problem dependent. So after applying the threshold value the **certain rules** we are left with are as follows:

If the tumor is large then D_A is present is 0.5
If the tumor is large and red then D_A is present is 0.5
If the tumor is large and blue then D_A is present is 0.5
If the tumor is red then D_B is present is 0.5
If the tumor is large and red then D_B is present is 0.6
If the tumor is small and red then D_B is present is 0.5
If the tumor is small and blue then D_B is present is 0.5

Next we find the possible rules by using the second function which is $J(X \# Y)$. Again we choose a threshold value and discard any rules which falls below this threshold value. These values measure the degree to which X intersects Y , and the rules generated by these are the possible rules. So carrying out the computations we get :

$$\begin{array}{lll}
J(S \# D_A) = .7 & J(S \# D_B) = .6 & J(R \# D_A) = .7 \\
J(R \# D_B) = .8 & J(B \# D_A) = .7 & J(S \# D_B) = .7 \\
J(L \cap R \# D_A) = .7 & J(L \cap R \# D_B) = .8 & J(L \cap B \# D_A) = .7
\end{array}$$

$$J(L \cap B \# D_B) = .7$$

$$J(S \cap R \# D_A) = .4$$

$$J(S \cap R \# D_B) = .5$$

$$J(S \cap B \# D_A) = .7$$

$$J(S \cap B \# D_B) = .7$$

For the possible rules let us set the threshold value (α) to .6. This would yield the following possible rules.

If the tumor is large then D_A is possible 0.7

If the tumor is small then D_A is possible 0.7

If the tumor is red then D_A is possible 0.7

If the tumor is blue then D_A is possible 0.7

If the tumor is large and red then D_A is possible 0.7

If the tumor is large and blue then D_A is possible 0.7

If the tumor is small and blue then D_A is possible 0.7

If the tumor is large then D_B is possible 0.8

If the tumor is small then D_B is possible 0.6

If the tumor is red then D_B is possible 0.8

If the tumor is blue then D_B is possible 0.7

If the tumor is large and red then D_B is possible 0.8

If the tumor is large and blue then D_B is possible 0.7

If the tumor is small and blue then D_B is possible 0.6

Now we are ready to extract the certain and possible rules. In the next section we look at how this is done.

4.3 Extracting Rules

The method used for extracting rules differ for the certain and possible rules. We will look at each case individually. First we look at how to extract certain rules. To extract certain rules:

- 1) All rules with unique degrees of belief are kept.

- 2) In case two or more rules have the same degree of belief then the one with the smaller number of attributes are kept.

Applying these rules to the above stated certain rules we get the following extracted certain rules.

If the tumor is large then D_A is present is 0.5

If the tumor is red then D_B is present is 0.5

If the tumor is large and red then D_B is present is 0.6

If the tumor is small and blue then D_B is present is 0.5

Now we see how to extract possible rules. The steps are as follows:

- 1) All rules with unique degrees of belief are kept.
- 2) In case two or more rules have the same degree of belief then the one with the larger number of attributes are kept.

Applying these rules to the possible rules shown above we get the following extracted possible rules:

If the tumor is large and red then D_A is possible 0.7

If the tumor is large and blue then D_A is possible 0.7

If the tumor is small and blue then D_A is possible 0.7

If the tumor is large and red then D_B is possible 0.8

If the tumor is large and blue then D_B is possible 0.7

If the tumor is small and blue then D_B is possible 0.6

4.4 Measuring Definability

Now we turn our attention to defining the fuzzy terms involved in the diagnosis as a function of the terms used in the symptoms. How well we are able to do this, is a function of how much the decision follows the conditions. The concepts explained here are from [2]. Let $\{B_i\}$ be a finite family of fuzzy sets which does not necessarily form a partition of the universal set. Let A be a fuzzy set. Then we can define the lower approximation of A through $\{B_i\}$ as

$$\underline{R}(A) = \bigcup_i (B_i \subset A) B_i.$$

In cases where $I(B_i \subset A)$ is less than some threshold α it is advantageous to throw away all the sets B_i . In this case we have

$$\underline{R}(A)_\alpha = \bigcup_i (B_i \subset A) B_i.$$

Similarly we show an upper approximation of A through $\{B_i\}$

$$\overline{R}(A) = \bigcup_i (B_i \# A) B_i$$

and

$$\overline{R}(A)_\alpha = \bigcup_i (B_i \# A) B_i$$

Returning to our initial example and applying these concepts: if we choose α to be .5 and $B_1 = L$; $B_2 = L \cap R$; $B_3 = L \cap B$; $B_4 = S \cap B$ then we have

$$\underline{R}(D_A)_{.5} = .5 L \cup .5(S \cap B)$$

Thus we can use a combination of Large and Small and Blue can be used to describe the set of patients that are certainly sick through the symptoms L,S,R,B. Similarly, if we pick α to be .6 then we get

$$\overline{R}(D_A)_{.6} = .7L \cup .7S \cup .7R \cup .7B$$

REFERENCES

- [1] Cohen, P.R. and Grinberg, M. R., A framework for heuristic reasoning about uncertainty, *Proc. 8th IJ CAI, Karlsruhe, W. Germany*, pp. 355 - 357 (1983).
- [2] de Korvin, Andre', Bourgeois, B., Kleye, R. *Extracting Fuzzy rules under Uncertainty and measuring definibility using rough sets*, Technical Report, University of Houston - Downtown, Houston, Tx.
- [3] Ehrenberg, ASC (1982) *A Primer in Data Reduction* John Wiley, Chichester.
- [4] Grzymala-Busse, J. W. "Knowledge acquisition under uncertainty - a rough set approach," *Journal of Intelligent and Robotic Systems* 1, pp. 3 -16 (1988).
- [5] Grzymala-Busse, J. W., On the reduction of knowledge representation systems, *Proc. 6th Internat. Workshop on Expert Systems and their Applications, Avignon, France*. Vol. 1, pp. 587- 599 (1986).
- [6] Johnson-Laird, PN and Wason, Pc (Eds) (1977) *Thinking : Readings in Cognitive Science* Cambridge University Press.
- [7] McCarthy, J., Circumscription - a form of non-monotonic reasoning, *AI* 13, 27-39 (1980).
- [8] McDermott, D. and Doyle, J., Non-monotonic logic I *AI* 13, 41-72 (1980).
- [9] Mamdani, A., Efstathiou, J., and Pang, D., Inference under uncertainty, *Expert Systems 85 . Proc. Fifth Tech. Conf. British Computer Soc., Specialist Group on Expert Systems*, pp. 181-194(1985).
- [10] Shafer, G., *Mathematical theory of Evidence*, Princeton University Press (1976).